Deep Learning - Week 3

Use the following data to answer the questions 1 to 2

A neural network contains an input layer $h_0 = x$, three hidden layers (h_1, h_2, h_3), and an output layer O. All the hidden layers use the Sigmoid activation function, and the output layer uses the Softmax activation function.

Suppose the input $x \in \mathbb{R}^{200}$, and all the hidden layers contain 10 neurons each. The output layer contains 4 neurons.

1. How many parameters (including biases) are there in the entire network?

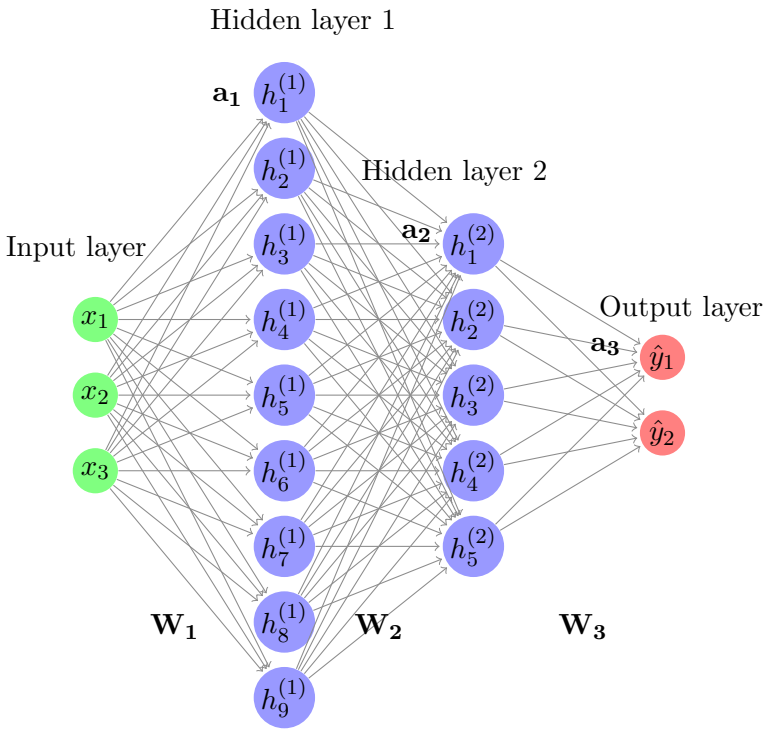
   Correct Answer: 2274
   Solution:
   Number of Parameters
   Input Layer to h_1: $200 \times 10 + 10 = 2010$
   h_1 to h_2: $10 \times 10 + 10 = 110$
   h_2 to h_3: $10 \times 10 + 10 = 110$
   h_3 to Output Layer: $10 \times 4 + 4 = 44$
   Total Parameters: $2010 + 110 + 110 + 44 = 2274$

2. Suppose all elements in the input vector are zero, and the corresponding true label is also 0. Further, suppose that all the parameters (weights and biases) are initialized to zero. What is the loss value if the cross-entropy loss function is used? Use the natural logarithm (ln).

   Correct Answer: Range(1.317,1.455)
   Solution:
   Loss with Zero Inputs and Parameters Input: $x = 0$, weights and biases $= 0$.
   Hidden Layers: $\sigma(0) = 0.5$.
   Output Layer Logits: $[0, 0, 0, 0]$.
   Softmax: $\text{Softmax}(z_i) = \frac{1}{4}$, $\forall i$.
   Cross-Entropy Loss: $-\ln\left(\frac{1}{4}\right) = \ln(4) \approx 1.386$.

Use the following data to answer the questions 3 to 4

The diagram below shows a neural network. The network contains two hidden layers and one output layer. The input to the network is a column vector $x \in \mathbb{R}^3$. The first hidden layer contains 9 neurons, the second hidden layer contains 5 neurons and the output layer contains 2 neurons. Each neuron in the l^{th} layer is connected to all the neurons in the $(l+1)^{th}$ layer. Each neuron has a bias connected to it (not explicitly shown in the figure).



In the diagram, $\mathbf{W_1}$ is a matrix and $\mathbf{x}, \mathbf{a_1}, \mathbf{h_1}$, and $\mathbf{O}$ are all column vectors. The notation $\mathbf{W_i}[\mathbf{j}, :]$ denotes the j^{th} row of the matrix $\mathbf{W_i}$, $\mathbf{W_i}[:, \mathbf{j}]$ denotes the j^{th} column of the matrix $\mathbf{W_i}$ and $\mathbf{W_{k_{ij}}}$ denotes an element at i^{th} row and j^{th} column of the matrix $\mathbf{W_k}$.

3. Choose the correct dimensions of $\mathbf{W_1}$ and $\mathbf{a_1}$

   (a) $\mathbf{W_1} \in \mathbb{R}^{3\times 9}$

   (b) $\mathbf{a_1} \in \mathbb{R}^{9\times 5}$

   (c) $\mathbf{W_1} \in \mathbb{R}^{9\times 3}$

   (d) $\mathbf{a_1} \in \mathbb{R}^{1\times 9}$

   (e) $\mathbf{W_1} \in \mathbb{R}^{1\times 9}$

   (f) $\mathbf{a_1} \in \mathbb{R}^{9\times 1}$

   Correct Answer: (c),(f)
   Solution:

4. How many learnable parameters(including bias) are there in the network?

   Correct Answer: 98
   Solution:
   Number of parameters in $\mathbf{W_1}$: $(\mathbf{9} * \mathbf{3}) + \mathbf{9}$
   Number of parameters in $\mathbf{W_1}$: $(\mathbf{5} * \mathbf{9}) + \mathbf{5}$
   Number of parameters in $\mathbf{W_1}$: $(\mathbf{2} * \mathbf{5}) + \mathbf{2}$
   Total: $36 + 50 + 12 = 98$.

5. We have a multi-classification problem that we decide to solve by training a feedforward neural network. What activation function should we use in the output layer to get the best results?

   (a) Logistic

   (b) Step function

   (c) Softmax

   (d) linear

   Correct Answer: (c)
   Solution: Softmax works best on multilayer classification problems since it is scale-invariant and outputs a probability distribution.

6. Which of the following statements about backpropagation is true?

   (a) It is used to compute the output of a neural network.

   (b) It is used to optimize the weights in a neural network.

   (c) It is used to initialize the weights in a neural network.

   (d) It is used to regularize the weights in a neural network.

   Correct Answer: (b)
   Solution: Backpropagation is a commonly used algorithm for optimizing the weights in a neural network. It works by computing the gradient of the loss function with respect to each weight in the network, and then using that gradient to update the weight in a way that minimizes the loss function.

7. Given two probability distributions p and q, under what conditions is the cross entropy between them minimized?

   (a) All the values in p are lower than corresponding values in q

   (b) All the values in p are higher than corresponding values in q

   (c) $p = 0$(0 is a vector)

   (d) $p = q$

   Correct Answer: (d)
   Solution:Cross entropy is lowest when both distributions are the same.

8. Given that the probability of Event A occurring is 0.18 and the probability of Event B occurring is 0.92, which of the following statements is correct?

   (a) Event A has a low information content

   (b) Event A has a high information content

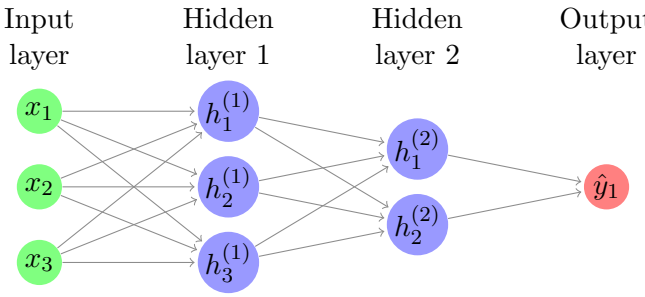
   (c) Event B has a low information content

(d) Event B has a high information content

Correct Answer: (b),(c)
Solution: Events with high probability have low information content while events with low probability have high information content.

Use the following data to answer the questions 9 and 10

The following diagram represents a neural network containing two hidden layers and one output layer. The input to the network is a column vector $x \in R^3$. The activation function used in hidden layers is sigmoid. The output layer doesn't contain any activation function and the loss used is squared error loss $(pred_y - true_y)^2$.



The following network doesn't contain any biases and the weights of the network are given below:

$$\mathbf{W_1} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & -1 & 1 \\ 1 & 2 & -2 \end{bmatrix} \quad \mathbf{W_2} = \begin{bmatrix} 1 & 1 & 2 \\ 3 & 1 & 1 \end{bmatrix} \quad \mathbf{W_3} = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

The input to the network is: $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

The target value y is: $\mathbf{y} = 5$

9. What is the predicted output for the given input x after doing the forward pass?

Correct Answer: Range(2.9,3.0)
Solution:
Doing the forward pass in the network we get

$$\mathbf{h_1} = \mathbf{W_1} \cdot \mathbf{x_1} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & -1 & 1 \\ 1 & 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ 3 \end{bmatrix}$$

$$\mathbf{a_1} = \text{sigmoid}(\mathbf{h_1}) = \begin{bmatrix} 0.997 \\ 0.731 \\ 0.952 \end{bmatrix}$$

$$\mathbf{h_2} = \mathbf{W_2} \cdot \mathbf{a_1} = \begin{bmatrix} 1 & 1 & 2 \\ 3 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.997 \\ 0.731 \\ 0.952 \end{bmatrix} = \begin{bmatrix} 3.632 \\ 4.674 \end{bmatrix}$$

$$\mathbf{a_2} = \text{sigmoid}(\mathbf{h_2}) = \begin{bmatrix} 0.974 \\ 0.990 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0.974 \\ 0.990 \end{bmatrix} = 2.954$$

10. Compute and enter the loss between the output generated by input $\mathbf{x}$ and the true output $\mathbf{y}$.

    Correct Answer: Range(3.97,4.39)
    Solution: Loss=$(5 - 2.954)^2 = 4.1861$